Class of Service based AS Interconnection

Th. M. Knoll, Chair of Communication Networks, Chemnitz University of Technology, Germany
Email: knoll@etit.tu-chemnitz.de

***Abstract*—The increasing number of delay and loss critical services in packet networks require differentiated packet handling in the forwarding plane. Quality of Service (QoS) guarantees can be given for networks using resource reservation and admission control. However, such strategies require complex control plane extensions and might lead to higher operation expenditures.**

Network operators therefore often use over-provisioning and traffic differentiation to offer cheaper Class of Service (CoS) quality in their internet protocol (IP) packet networks.

The number of differentiated classes and their autonomous system (AS) internal implementation is at the operator's choice.

This paper proposes a signalling concept for inter-AS layer three Class Set signalling, supported classes, their encoding and packet rate limitations. It makes use of the Border Gateway Protocol (BGP) as the predominantly used routing protocol for AS peering communication. The paper specifies two new non-transitive attributes, which enable adjacent peers to signal Class of Service capabilities and admission control limitations. The new "CoS Capability Attribute" and the "CoS Parameter Attribute" are simple data structures, which signal the classes, their per hop behaviour (PHB) ID code and the token bucket control performed at the ingress AS border router for rate limitation purposes. The denoted Class of Service forwarding support is meant as the AS externally available (transit) Class of Service support.

The approach is now work in progress at the IETF.

***Index Terms*—BGP, QoS, class set signalling, inter-AS CoS, CoS capability attribute, CoS parameter attribute**

I. Introduction

Quality of Service (QoS) can be achieved using either resource reservation with admission control or through service differentiation based on prioritized traffic classes. This paper focuses on traffic class priority only and will support coarse QoS in terms of "Class of Service (CoS)". The Differentiated Services (DiffServ) architecture [4] is the layer 3 (L3) priority mechanism used today. In the architecture, per domain behaviours (PDBs) are constructed by means of the creation of traffic aggregates by applying rules at the ingress and associating those aggregates with certain path forwarding treatments – per hop behaviours (PHBs). Packets belonging to the same aggregate are carrying the same differentiated services codepoint (DSCP) in their IP header.

Per domain behaviours are constructed as quantifiable forwarding behaviour in a Differentiated Services network domain based on per hop behaviours in relaying nodes. Due to the lack of standardized PDB and the not targeted quantification of forwarding parameters in this approach, the paper focuses on defined PHB as forwarding behaviours at the interconnection points. The inter-AS signalling is also based on PHB IDs in conformance to RFC 3140 [3].

Autonomous System (AS) operators can freely choose and configure the set of supported Classes of Service provided for transit traffic across their network. These AS-internal CoS policy decisions are made independently and will not necessarily be shared or synchronized with neighbouring ASes. This paper proposes a modified Border Gateway Protocol (BGP – see section V) QoS signalling mechanism, which provides a consistent unidirectional inter-AS CoS information exchange.

The BGP transferred CoS information enables peering partners to adopt their forwarded traffic at the exchange point to the supported transit CoS of the downstream neighbour. Forming traffic aggregates and possibly performing traffic shaping is now controllable by the upstream AS. The disclosed token bucket rate limitation at the downstream AS ingress is known to both partners and provides a fair and square base of interconnection. It also protects the downstream AS from excessive overload on certain classes. The ingress reaction (remarking or dropping) to excess traffic is also signalled by the defined attributes.

QoS in this approach refers to primitive traffic separation into several classes, which will experience differently prioritized forwarding behaviour in relaying nodes. No QoS parameters are guaranteed, but enqueueing in separate forwarding queues is aspired, which leads to a better than best effort forwarding behaviour. The approach is currently implemented for experimental analysis in the so called Quagga routing suite under the Linux operating system. Results will be published as they become available.

The word "peering" – as opposed to "transit" – relates to free of charge AS interconnection. Since pricing is not addressed in the paper, the word is used as general term for the interconnection of ASes.

This paper first outlines a quick overview of the DiffServ architecture and their inter-domain PHB ID encoding. It highlights the flexible class selection options and focuses on four groups of PHBs. The major contribution of this work is

Manuscript received November 19, 2008.

Th. M. Knoll is with the Faculty of Electrical Engineering and Information Technology, Chemnitz University of Technology, 09107 Chemnitz, Germany (phone: +49 (0) 371 531 33246; fax: +49 (0) 371 531 833246; e-mail: knoll@etit.tu-chemnitz.de).

described in section V. BGP is used as signalling transport mechanism for the two new BGP attributes (CoS Capability Attribute and CoS Parameter Attribute). Section V gives a short overview on BGP operation and the BGP message structure before the actual attribute definition follows.

The last five sections explain the proposed signalling on an example network, outline its usage, name related work and consider security, confidentiality and business aspects.

II. Related work

A number of QoS improvement approaches have been proposed before and a selection will be briefly mentioned in this section. Most of the approaches perform detailed QoS parameter signalling.

[8] defines the QOS_NLRI attribute, which is used for propagating QoS-related information associated to the NLRI information conveyed in a BGP UPDATE message. Single so called "QoS routes" are signalled, which fulfil certain QoS requirements. Several information types are defined for the attribute, which concentrate on rate and delay type parameters.

[6] is based on the specified QOS_NLRI attribute and introduces some modifications to it. The notion of AS-local and extended QoS classes is used, which effectively describes the local set of QoS performance parameters or their cross-domain combined result. Two groups of QoS delivery services are distinguished, where the second group concentrates on ID associated QoS parameter propagation between adjacent peers. The first group is of more interest for this paper since it concentrates on the "identifier propagation" - such as the DSCP value for example. This signalling is specified for the information exchange between adjacent peers and assumes the existence of extended QoS classes and offline traffic engineering functions.

Another approach is described in [5]. It associates a list of QoS metrics with each prefix by extending the existing BGP AS_PATH attribute format. Hop-by-hop metric accumulation is performed as the AS_PATH gets extended in relaying ASes. Metrics are generically specified as a list of type length value (TLV) style attribute elements. The metrics such as bandwidth and delay are exemplary mentioned in the draft.

One contribution specialized in the signalling of Type Of Service (TOS) values which are in turn directly mapped to DSCP values in section 3.2 of the draft [21]. The TOS value is signalled within an Extended Community Attribute and, if it is understood correctly, will be applied to a certain route. An additional value field is used to identify, which routes belong to which signalled TOS community. Who advertises such attributes and whether they are of transitive or non-transitive type remains unspecified.

A comprehensive analysis is given in [1]. This "Inter-provider Quality of Service" white paper examines the inter-domain QoS requirements and derives a comprehensive approach for the introduction of at least one QoS class with guaranteed delay parameters. The implementation aspects of metering, monitoring, parameter feedback and impairment allocations are all considered in the white paper. However, QoS guarantees and parameter signalling is beyond the intention of this paper.

A very extensive work has been published in [16]. It goes far beyond this limited CoS approach of this paper. The so called “loose guarantees solution” in that work is one offered option that also renounces end-to-end QoS guarantees. However, it still performs mutual negotiations on performance parameters and bandwidth requirements.

Other documents may also be considered as related work as long as they convey QoS marking information, that might be "misused" for CoS signalling.

One example is the usage of the "Traffic Engineering Attribute" as defined in IETF draft [18]. However, the attribute is non-transitive and the LSP encoding types are not generally applicable to inter-domain peering types. Its usage of the targeted QoS marking signalling is not possible.

The second example is the current "Dissemination of flow specification rules" draft [15]. It defines a new BGP NLRI encoding format, which can be used to distribute traffic flow specifications. Such flow specification can also include DSCP values as type 11 in the NLRI. Furthermore, one could signal configuration actions together with the DSCP encoding, which could be used for filtering purposes or even trigger remarking and route selection with it. Such usage is not defined in the draft and can hardly be achieved because of the following reason. The flow specification is focused on single flows, which might even be part of an aggregate. Such fine grained specification is counterproductive for this coarse grained general CoS capability approach.

The proposed approach of this “Class of Service based inter-AS Peering” paper is – in an earlier stage – work in progress at the IETF [14].

The transitive signalling of supported classes and their markings between ASes is addressed in a separate but complementary IETF draft [13], which is not described in this paper. The PHB IDs, that are used in both drafts allow for the combined usage of globally visible Class of Service markings and their locally applied token bucket filtering.

The advantage of this “better than best effort” approach, as compared to the listed related work, lies in its simplicity and free to join nature. End-to-End QoS is not achieved, but traffic separation on interconnection points is provided. The missing quality guarantees and the associated service level agreements on QoS with individually agreed on classes, parameters, measurement procedures and fines are key for the acceptance and quick global deployment of the concept.

III. Differentiated Services Architecture

The DiffServ architecture is defined in RFC2475 [4] and RFC2474 [17]. “Differentiated Services” in its broader sense

encompasses significant quantitative or statistical characteristics of packet transmission in one direction across a set of one or more paths within a network or simply provide some relative priority of access to network resources. The latter will be focused on in the paper.

Service differentiation is predominantly looked at from the technical perspective. However, pricing differentiation is a possible consequence.

DiffServ is a Quality of Service provisioning concept within a network domain that applies rules at the edges to create traffic aggregates and couples each of these with a specific forwarding path treatment in the domain by means of a Differentiated Services CodePoint (DSCP) in the IP header. A network domain, that supports the Differentiated Services forwarding behaviour concept is called a "DiffServ Domain (DS)".

A. DSCP encoding

As shown in Fig. 1, the Class of Service encoding follows a two step approach. First the aspired CoS is expressed as Per Hop Behaviour, which is then represented by one of the locally applied DSCP values for that PHB. There is no limitation for available per hop behaviours and future services might require to continuously add new PHB definitions as they arise. The DSCP encoding, however, is limited to a 6 bit structure, which allows for 64 locally distinguished DiffServ traffic aggregates (see Fig. 2). Currently there are about 20 DSCP values commonly used.

Some DSCP values are fixed for certain PHBs, but the majority can freely be chosen by operators for local PHB selection.

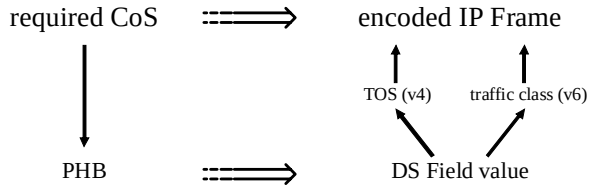

Fig. 1 CoS to DSCP marking

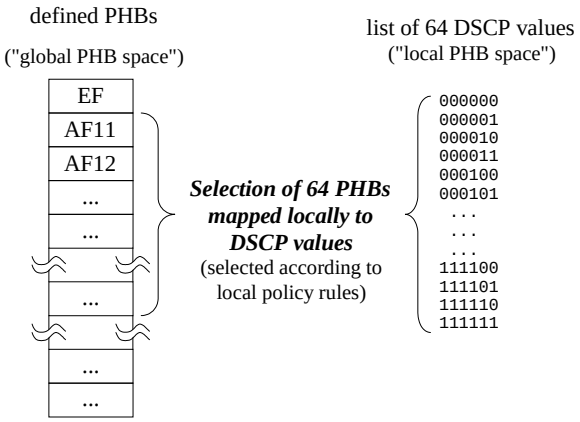

Fig. 2 PHB mapping

The Differentiated Services CodePoint values are encoded in the so called DS field of the IP header. However, this field is a redefinition of the original IP version 4 Type of Service (TOS) field or the IP version 6 Traffic Class field, see Fig. 3.

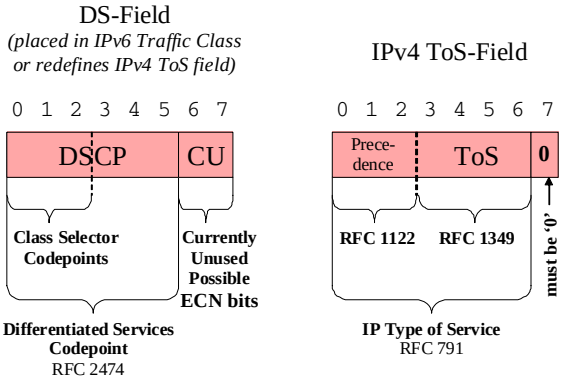

Fig. 3 The DS field in the IP header

This redefinition led to the common approach to support the original IP precedence (bits 0, 1 and 2) definition as so called "Class Selector Codepoints" in the new DSCP value space.

Other fixed DSCP values result from certain PHB specifications.

1) "Best Effort (BE)" PHB

Although the IP header specification included the aforementioned precedence and type of service indications, many software implementations and router systems ignored those bits and assumed a value of zero in this field for ordinary IP packets. That is why the DSCP value of '000000' is traditionally set for the usual "Best Effort" forwarding behaviour.

2) "Expedited Forwarding (EF)" PHB

RFC3246 [10] defines the so called "Expedited Forwarding" PHB, which is intended to provide a building block for low delay, low jitter and low loss services by ensuring that the EF aggregate is served at a certain configured rate. RFC3247 [8] gives supplemental information on EF as well as implementation examples.

The DSCP for the EF PHB is defined to '101110'.

3) "Assured Forwarding (AF)" PHB

The "Assured Forwarding" PHB is defined as PHB group in RFC 2597 [11]. As RFC 3260 states: "Assured Forwarding (AF) is a type of forwarding behaviour with some assigned level of queuing resources and three drop precedences. An AF PHB Group consists of three PHBs, and uses three Diffserv Codepoints (DSCPs). RFC 2597 defines twelve DSCPs, corresponding to four independent AF classes. The AF classes are referred to as AF1x, AF2x, AF3x, and AF4x (where 'x' is 1, 2, or 3 to represent drop precedence). Each AF class is one instance of an AF PHB Group."

The DSCP values for the AF classes are shown in Table I.

Table I
DSCP values for AF PHB classes and drop precedence within each class

AF class	DP low	DP medium	DP high
1	**001**010	**001**100	**001**110
2	**010**010	**010**100	**010**110
3	**011**010	**011**100	**011**110
4	**100**010	**100**100	**100**110

4) "Lower Effort (LE)" PHB

The "Lower Effort (LE)" PHB is defined in RFC 3662 [5]. It is intended for traffic of sufficiently low value (where "value" may be interpreted in any useful way by the network operator), in which all other traffic takes precedence over LE traffic in consumption of network link bandwidth. This PHB is well suited for "optional" traffic, which might or might not be forwarded.

There is no fixed DSCP value assigned for LE within the RFC. It rather suggests to use either an experimental DSCP or an AF DSCP or the Class Selector 1 (DSCP='001000').

5) "Voice" PHB

The IETF transport working group is proposing a separate DSCP value for connection admission controlled voice traffic. Forwarding voice traffic as EF PHB does not allow to distinguish between "ordinary" voice calls and those that cooperate with network control. The latter should be preferred before the other and could be used for emergency services etc.

The work in [2] has not reached RFC status and no DSCP value has been reserved for this PHB as yet.

However, as soon as such emergency calls can be distinguished within a network, the inter-AS peering should be able to differentiate this traffic as well.

B. Inter-domain PHB ID codes

The encoding mechanism for the identification of differentiated services PHB in protocol messages is defined in RFC 3140 [3]. It addresses the difficulty of consistent PHB identification under the circumstances of locally mapped DSCP values to PHBs. Inter-domain PHB signalling in particular needs a reliable encoding mechanism besides the 6 bit IP header DS field. RFC 3140 therefore defines an unsigned 16 bit binary encoding to uniquely identify PHBs and/or sets of PHBs.

PHB identification codes for standards track PHBs with assigned DSCP values follow the structure in Fig. 4.

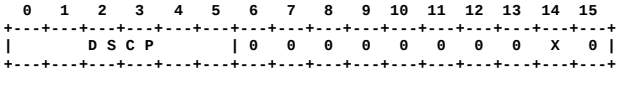

Fig. 4 PHB ID codes for standards track PHBs

Bit 14 of this encoding structure indicates, whether the a single PHB ('0') or a group of PHBs ('1') is addressed.

C. Selected PHB groups

The mentioned PHB groups in section A are a limited subset of the generally unlimited PHB specification space. Current inter-AS peering only offers the BE PHB as the one and only traffic class.

This concept, however, introduces a coarse differentiation of traffic aggregates as a trade-off between CoS based forwarding at the peering exchanges and the necessity of simple and stable peering conditions. Therefore, only a small selection of supported PHBs should be signalled and used at the inter-AS peering points.

As **guideline to operators**, this paper suggests the following two Class Set options in the order of preference.

The first collection of "**Basic CoS**" would comprise the **EF** PHB for delay critical services, one or more **AF** classes for "higher value" traffic with relative scheduling and dropping precedence, the **BE** PHB for the "normal" Internet traffic and the **LE** PHB for "optional" (background) traffic.

Voice traffic, as a very important and delay sensitive traffic type, is currently mapped into EF and will keep this assignment.

The resulting four class concept is considered a sufficiently fine grained traffic differentiation compared to the current "BE only" peering.

As second option, a simple two class concept might arise, that allows for the distinction between "normal" Internet traffic and "optional" (background) traffic only.

IV. Cross-Domain and Inter-AS CoS

Three different problems are identified for consistent end-to-end QoS handling. Two are cross-domain QoS signalling and cross-layer traffic class mapping, which have been addressed in an IETF draft [13] outside of this paper's scope.

The third considers the local tuning of classes of service at the peering point based on the supported CoS at either side. Local (non-transitive) signalling is used in order to indicate the respective CoS support of the downstream AS for transit traffic. Furthermore, the described concept predetermines DSCP values for the respective PHBs, where both concept supporters adopt their local mapping in order to meet a common DSCP marking style at the peering point.

Furthermore, a major advantage of this concept is the additional parameter exchange about the applied ingress token bucket rate limitation. This mechanism is in place in order to prevent overload situations in higher value classes. Each supported PHB group can be associated with a token bucket parameter set in order to indicate to the peering partner the measurement method and the resulting consequences for excess traffic in a fair and square manner.

This way, remarking, dropping and possibly accounting figures can be forecasted and reproduced by the relaying AS. The usage of egress traffic shapers based on the signalled parameter set is out of scope for this paper.

V. MODIFIED BGP QOS SIGNALLING

The Border Gateway Protocol version 4 (BGP-4) is the predominant inter-AS routing protocol. It is the base for the proposed new CoS signalling mechanism of this paper and will therefore be explained in more detail in the following paragraphs.

A. Inter-AS routing and signalling using BGP

IP networking between networks of different providers is realized by peering points and associated peering agreements called "Service Level Agreement (SLA)".

The BGP-4 is used for the mutual information exchange about reachable IP networks.

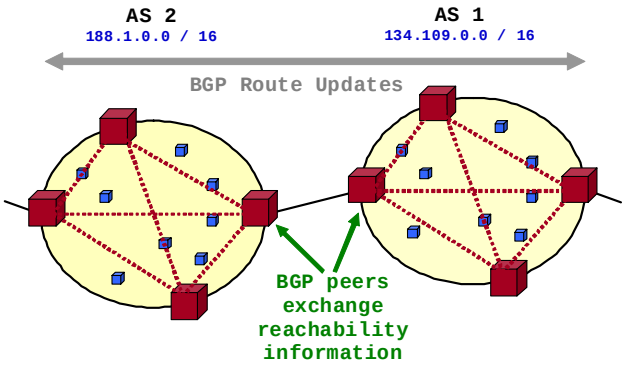

Fig. 5 BGP peering between neighbouring ASes

Fig. 5 shows a simple network setup where AS border routers establish BGP peering sessions between configured neighbouring peers and send BGP messages for a secured and reliable global routing information exchange.

The virtually fully meshed BGP peer overlay network allows for a controlled inter-AS information exchange, which can be used to signal reachability and other information between all existing AS border routers in a unique and consistent manner.

The core BGP routing information messages are so called BGP UPDATE messages, which consist of four parts:

1. Message header,
2. Withdrawn routes,
3. Path attributes and
4. Network Layer Reachability Information – NLRI.

The UPDATE messages include BGP Path Attributes, which signal origin and routing details. This information can be used to further control the route filtering and the advertisement process.

These attributes are of central concern for this paper and are therefore explained in detail.

BGP Path Attributes

A number of BGP path attributes are defined and can be grouped into "well-known vs. optional" (see Fig. 6) and "transitive vs. non-transitive" attributes.

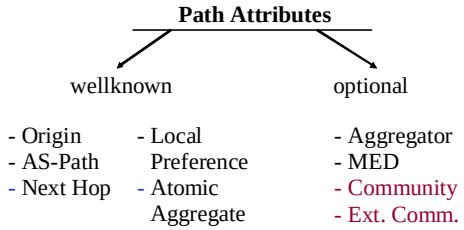

Fig. 6 BGP path attributes [7], [11], [19] and [20]

This paper focuses on the use of so called "Extended Community Attributes", which are defined in RFC4360 [20] as well as defines an additional new optional non-transitive attribute for parameter signalling.

Community attributes (including Extended Community attributes) are not essential for BGP peering and proper interworking of inter-AS IP routing. They are optional attributes and can be marked as transitive (signalled across all ASes) or non-transitive (signalled only between neighbouring AS peers).

B. Definition and Usage of the CoS Capability Attribute

A new Extended Community attribute is defined here, which carries CoS capability information as binary bit encoding.

This new BGP Extended Community attribute is called "**CoS Capability Attribute**" (Fig. 7). Each supported PHB group is indicated as single CoS Flag.

The CoS Capability Attribute is a non transitive optional BGP attribute, with the Type Code 16. IANA has assigned the type code 0x40 for this Extended Community Attribute [12].

Fig. 7 depicts the attribute structure.

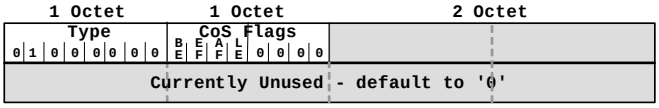

Fig. 7 CoS Capability Attribute structure

CoS Flags:

All flags – but BE – default to a value of '0'.

Table II shows the bit encoding of the flags field.

Table II CoS Capability Attribute – CoS Flags field

Bit	Flag	Encoding
0	BE	Default to '1' to signal general "Best Effort" PHB support
1	EF	'1' … "Expedited Forwarding" PHB support [10]
2	AF	'1' … "Assured Forwarding" PHB group support [11]
3	LE	'1' … "Lower Effort" PHB support [5]
4 .. 7	unused	Default to '0'

This encoding signals the BE, EF, AF group and LE support of the respective advertising AS. The implied Per

Hop Behaviour Identification Codes follow the definition as standardized in [3]. An AS is regarded as “AF capable” as soon as at least one of the available AF1x, AF2x, AF3x and AF4x PHB groups is signalled to be supported.

A common PHB encoding is documented in Fig. 8, which ensures consistent interpretation of the exchanged signalling information.

PHB ID encoding of the signalled PHBs:

The CoS Flag field refers to certain PDB definitions, which are taken from the respective specifications and are fixed in their encoding values for the usage within this CoS-based inter-AS peering concept.

```
BE:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   0   0   0   0   0 | 0   0   0   0   0   0   0   0   0   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

EF:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 1   0   1   1   1   0 | 0   0   0   0   0   0   0   0   0   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

AF1x:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   0   1   0   1   0 | 0   0   0   0   0   0   0   0   1   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

AF2x:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   1   0   0   1   0 | 0   0   0   0   0   0   0   0   1   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

AF3x:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   1   1   0   1   0 | 0   0   0   0   0   0   0   0   1   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

AF4x:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 1   0   0   0   1   0 | 0   0   0   0   0   0   0   0   1   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

LE:
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| 0   0   1   0   0   0 | 0   0   0   0   0   0   0   0   0   0 |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Fig. 8 PHB ID encoding of supported PDB

The CoS Capability Attribute is used as primitive means to signal the general availability of the supported PHBs in the

neighbouring AS. The attribute is included within the attribute section of an BGP UPDATE message and is therefore associated to the NLRI information of the same message.

C. Definition and Usage of the CoS Parameter Attribute

A second and entirely new defined path attribute is used to signal token bucket parameter sets together with the supported classes of service.

The so called “CoS Parameter Attribute” is an optional non-transitive BGP attribute of variable length.

The attribute contains the PHB ID code, a flags field and the set of token bucket parameters for each indicated PHB.

Fig. 9 depicts the attribute structure.

The actual length in byte of the attribute is indicated within the path attribute section of the UPDATE message and results for the number of token bucket associated PHBs times 24.

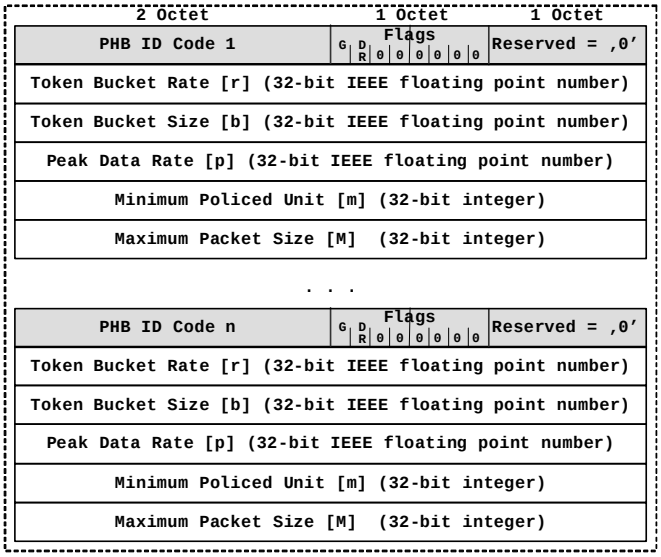

Fig. 9 CoS Capability Attribute structure

PHB ID Code:

The signalling of CoS parameters of a supported PHB groups reuses the definition from section III.B.

Flags:

Only two flags are defined. The resulting bits default to '0' and must be ignored on reception.

The 'G' flag signals, whether the limitations have global scope on all incoming traffic ('1') or are associated to traffic that is destined to destinations within the NLRI of the UPDATE message ('0'). NLRI specific limitation will supersede globally signalled ones for traffic destined to those NLRI destinations.

The 'DR' flag signals the applied handling of non-confirming traffic. DR='0' signals strict dropping of excess traffic. DR='1' signals the performed remarking of excess traffic packets to Best Effort traffic marking.

Traffic of type LE will not normally be associated with token bucket parameters and if so, will always being dropped in the excess case.

Token Bucket Parameter Set:

The definition of the signalled token bucket parameters follows the specification in RFC 2215 [21].

All rates are given in Bytes/s. Units and Size are expressed in Bytes.

VI. USAGE OF THE BGP COS CAPABILITY ATTRIBUTE

Providers may choose to analyze the neighbour’s CoS Capability Attributes and adopt the priority encoding according to their local policy. As soon as the mutual support

of a given Class of Service is signalled between both peering partners, they can rely on the forwarded DSCP marking with the forwarded data packets for the given mutually signalled PHB ID. Costly multi-layer ingress classification, if in place, may be omitted.

Whether the signalled CoS support may also lead to different IGP routing decisions or even effect BGP update filters is out of scope for this paper.

The associated Token Bucket parameters are to be used by the downstream AS at the ingress border router in order to prevent traffic class overload by non-conforming upstream neighbours. In the case of layer two internet exchange points, where several peering partners may send data traffic onto the same layer two output port of a switch, the sending upstream neighbour needs to be identified by means of its local MAC address as the sourcing address at the peering point.

The usage of the advertised Token Bucket parameters by the upstream AS at the egress border router is out of scope of this paper. Traffic shaping and excess punishment prediction could, however, be performed at operator's choice.

Since both attributes are non-transitive and of local scope, no judgement on the overall CoS support along an AS chained forwarding path can be made. This cross-domain information can only be revealed by means of the QoS Marking concept [13]. Close QoS treatment approximation across ASes and across several networking layers within the forwarding path can then be achieved using the CoS Capability and Parameter BGP attributes.

Frequent signalling of parameters within BGP UPDATE messages is of no use and even counterproductive for BGP stability. The proposed concept of Token Bucket parameter signalling for supported Class of Service is therefore limited to long range variations on rates and changes of the CoS sets. Changes in the signalled information are expected to happen on a daily, weekly or even longer time range.

VII. Impact on Routing and Forwarding

BGP route updates using the proposed attributes will be able to signal different CoS encodings and rate limitation between AS boundaries (eBGP) as well as within ASes (iBGP). Network operators are able to adjust their internal marking and route advertisement strategy as well as traffic engineering based on the signalled information. This internal strategy can transparently being reverted at the AS egresses to the advertised transit CoS encoding.

This mechanism eliminates the stringent BE peering constraint and enables a fair and square AS interconnection.

The experienced forwarding behaviour of individual IP packets will therefore be adopted to the targeted Class of Service and will be different in many cases to the currently experience forwarding treatment. However, this only influences scheduling and dropping priorities and not the path along which packets are forwarded.

Routing decisions, prefix aggregation and such are not touched be this concept. However, future extensions are not precluded per se, which might make use of the signalled CoS support for modified best path selection or path selection in a multi-path peering setup.

Both approaches are hardly used in practice and will not be available globally in the foreseeable future.

One major concern for the BGP operation is the size of the routing tables. About 250000 IP prefixes are currently stored and looked up in this routing information base, which places a strong burden on memory and processing power for BGP peers. Since IP prefix aggregation remains untouched by this Inter-AS CoS concept, BGP scalability and stability is not influenced detrimentally.

Triggering internal setups and update filters in the BGP route-maps will induce configuration work for the operator. The gained CoS forwarding support will, however, make up for it. CoS routing in the multi-path case is expected to raise major changes in the interconnection setups and needs further investigation for traffic churn and stability issues.

The signalling of CoS information does not lead to significantly increased routing update traffic. All route prefixes within a BGP UPDATE message are associated with the included CoS Capability Attribute Sets of the message, which is expected to add about 100 Bytes to an UPDATE message. The storage and processing is interface-local.

VIII. Impact on Security and Confidentiality

The proposed CoS Capability Attribute does not raise extra security concerns. Existing BGP security measures are in place through the reused transport in BGP UPDATE messages.

The disclosure of confidential network intrinsic information is of no concern since network operators have full control on which encoding is signalled and for which IP prefixes.

Furthermore, if the network internal QoS mechanisms and markings shall not be disclosed to BGP peers, but the proposed BGP CoS signalling shall still be supported, it is also feasible to signal a "faked" set of Classes of Service. This strategy of hiding actual implementations behind a generalized CoS set must be accompanied by appropriate translation and remarking functions at the advertising AS border routers.

IX. Business implications

Making AS-internally available traffic separations known globally, can easily lead to a traffic overload in a certain forwarding class. In order to ensure the proportional usage of the available options, this approach offers the rate limitation feature, which could possibly be aligned with a locally applied class-based accounting. That is, inter-AS traffic will be counted and priced according to the used Class of Service.

However, no complex service level agreements on QoS parameter boundaries and contractual penalties need to be setup for this limited CoS level of traffic separation. Especially in the two class setup with BE and LE PHBs, it is feasible to offer the LE acceptance as add on and market advantage.

REFERENCES

[1] Amante, S., Bitar, N., Bjorkman, N., and others, "Inter-provider Quality of Service - White paper draft 1.1", [Online]. Available: http://cfp.mit.edu/docs/interprovider-qos-nov2006.pdf

[2] Baker, F., Polk, J., Dolly, M., “An EF DSCP for Capacity-Admitted Traffic” September 2006. Available: http://tools.ietf.org/html/draft-baker-tsvwg-admitted-voice-dscp-01

[3] Black, D., Brim, S., Carpenter, B., and F. Le Faucheur, "Per Hop Behavior Identification Codes", RFC 3140, June 2001.

[4] Blake, S., Black, D., Carlson, M., Davies, D., Wang, Z. and W. Weiss, “An Architecture for Differentiated Services”, RFC 2475, December 1998

[5] Bless, R., Nichols, K., Wehrle, K., „A Lower Effort Per-Domain Behavior (PDB) for Differentiated Services“, RFC 3662, December 2003

[6] Boucadair, M., "QoS-Enhanced Border Gateway Protocol", IETF draft-boucadair-qos-bgp-spec-01 (work in progress), July 2005.

[7] Chandra, R., Traina, P., and T. Li, "BGP Communities Attribute", RFC 1997, August 1996.

[8] Charny, A., Bennett, J.C.R., Benson, K., et. al., ” Supplemental Information for the New Definition of the EF PHB (Expedited Forwarding Per-Hop Behavior)”, RFC 3247, March 2002

[9] Cristallo, G., "The BGP QOS_NLRI Attribute", IETF draft-jacquenet-bgp-qos-00 (work in progress), February 2004.

[10] Davie, B., Charny, A., Bennett, J.C.R., et. al., ”An Expedited Forwarding PHB (Per-Hop Behavior)”, RFC 3246, March 2002

[11] Heinanen, J. et. al., "Assured Forwarding PHB Group". RFC2597, June 1999

[12] IANA, “BGP Extended Communities Types”, IANA Protocol Registries [Online]. Available: http://www.iana.org/assignments/bgp-extended-communities

[13] Knoll, T., "BGP Extended Community Attribute for QoS Marking", IETF draft-knoll-idr-qos-attribute-02 (work in progress), July 2008.

[14] Knoll, T., "BGP Class of Service Interconnection", IETF draft-knoll-idr-cos-interconnect-00 (work in progress), July 2008.

[15] Marques, P., Sheth, N., Raszuk, R., Greene, B., and D. McPherson, "Dissemination of flow specification rules", IETF draft-ietf-idr-flow-spec-01 (work in progress), April 2008.

[16] Morand, P., Boucadair, M., Asgari, H., Egan, et al., “D1.4: Issues in MESCAL Inter-Domain QoS Delivery: Technologies, Bi-directionality, Inter-operability, and Financial Settlements”, MESCAL Consortium, January 2004. Available: http://www.ist-mescal.org/deliverables/MESCAL-D14-final-v2.pdf

[17] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[18] Ould-Brahim, H., "Traffic Engineering Attribute", draft-ietf-softwire-bgp-te-attribute-00 (work in progress), January 2008.

[19] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[20] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006.

[21] Schenker, S., Wroclawski, J., “General Characterization Parameters for Integrated Service Network Elements”, RFC 2215, September 1997.

[22] Zhang, Z., "ExtCommunity map and carry TOS value of IP header", IETF draft-zhang-idr-bgp-extcommunity-qos-00 (work in progress), November 2005.